Phoneme detection in resyllabified words

Vroomen, J.; de Gelder, B.

Published in:
Journal of the Acoustical Society of America

Publication date:
1996

9:00

5aPP5. Is impaired auditory temporal processing a cause of speech-language disorders? Negative evidence from psychoacoustic investigations. Charles S. Watson and Betty U. Watson (Dept. of Speech and Hearing Sci., Indiana Univ., Bloomington, IN 47405)

Psychoacoustic studies cast some doubt on the hypothesis that deficits in auditory temporal processing cause speech and language disorders. First, the temporal acuity of the auditory system so greatly exceeds that required by tasks considered to support this hypothesis [e.g., Tallal and Piercy, Nature 241, 468–469 (1973)] that disordered subjects would be quite unable to understand speech if their basic temporal processing were as imprecise as it is proposed to be. Second, in a study of 127 normal-hearing listeners, nonsense-syllable recognition was not correlated with temporal discrimination measured with various nonspeech stimuli, and the same result was later obtained for sentences, words and CV's, although all speech measures load on a common factor. Third, structural-equation models were fitted to data collected from 24 college students with specific reading disabilities, and 70 who were "normal." Data to be predicted included passage comprehension, word-attack skills, and word identification. A model using measures of speech perception and of intelligence yielded $0.7 > r^2 > 0.9$, but these predictions were not improved by nonspeech measures of auditory temporal processing. A viable hypothesis is that sensory processing speed may be only indirectly related to some speech-language disabilities, but more directly related to intelligence. [Work supported by NIH/NIDCD.]

9:15

5aPP6. Phoneme recognition by a non-native English-speaking cochlear implantee using the WSP, MSP, and spectra speech processors: A case study. Kathryn S. Copmann (Dept. of Speech Lang. Pathol. /Audiol., Loyola College, 4501 N. Charles St., Baltimore, MD 21210-2699)

This study presents the performance of a non-native English-speaking adult using the Nucleus 22 Channel cochlear implant. Included are the subject's (1) history, (2) promontory stimulation results, (3) audiological information, (4) pre- and post-implantation speech recognition test results, (5) post-implantation auditory phoneme recognition using the wearable speech processor (WSP), the minispeech processor (MSP), and the Spectra processor, and (6) the subject's subjective reports of benefits in terms of communication and social interactions. The subject, a highly motivated and intelligent male who, presented with a profound bilateral, progressive sensorineural hearing loss of unknown etiology, was 59 years old when the right ear was implanted. Prior to implantation the subject was aided in the left ear, but had never been aided in the right ear. The subject's performance using each of three processors is presented for medial vowel recognition and medial consonant recognition using auditory stimulation only. Results are displayed through confusion matrices generated based on the subject's responses to these acoustic stimuli. In summary, the vowel and consonant phoneme recognition of a subject with the Nucleus cochlear implant using various processors is presented. Improved phoneme recognition is demonstrated with filter banking versus feature extraction coding strategies.

9:30

5aPP7. A handicapped listener's articulation tests underwater. Akiteru Fukuda and Heiji Okada (Dept. of Commun., Shibaura Inst. of Tech., 3-9-14 Minato-ku Tokyo, 108 Japan)

There have been a number of papers in underwater sound propagation for a diver's mutual communication. A healthy man's monosyllable hearing test underwater has been reported previously. Now, a report of results on disabled (hard of hearing) listeners test is given. Sound propagation tests used a standard Japanese 25 m(L)×13 m(W)×1.5 m(D) waterpool. An underwater speaker Altec type UW-30 is used and is driven by a 10-W amplifier. A listener's monosyllable test signal is used with a standard voice number (101) by an NHK announcer. The distances between the radiation speaker and the disabled listener are agreed upon at 5, 10, and 20 m, and depths are changed from 1.3 to ~1.5 m. Radiator and listener's depths are held at the same level through every test. Depth simulation used with the air chamber at 10 kg/cm^2 normal air at Saitama Medical Univ., where pressures are controlled equivalent to 0-, 10-, 20-, and 30-m water depths. In the results of the articulation test, about the same scores are observed between the handicapped person and the healthy listener. Especially at the distance of 10-m, a handicapped person's average score is greater than 4% compared with the healthy person's score. Such a result requires further examinations.

9:45

5aPP8. Learning and generalization in auditory backward masking. Beverly A. Wright, Paul A. Johnston, and Miriam D. Reid (Keck Ctr., Box 0732, Univ. of California, San Francisco, CA 94143-0732)

Excessive auditory backward masking has been observed in elderly persons and in children with language impairments. The purposes of this project were to determine (1) whether practice can improve the ability to detect a backward-masked signal and, if so, (2) whether this improvement generalizes to other masking tasks. These issues were examined in six normal-hearing adults using an adaptive, 2IFC procedure. Each subject was asked to detect a tonal signal (10 ms, 1000 Hz) presented immediately before a noise masker (300 ms, 200–1800 Hz, 40-dB SPL spectrum level), on 900 trials per day for ten days. Before and after the training period, subjects were tested on five other masking tasks in addition to the trained task. A one-way ANOVA revealed a significant learning curve over the ten days of training [$F(9,45)=5.1$, $p<0.001$], with the average signal threshold decreasing from 63 to 53 dB SPL. Comparison of the pre- and post-test results showed significant threshold reductions in other backward-masking tasks, but not in simultaneous- and forward-masking tasks. These results indicate that (1) the interference from an auditory backward masker can be reduced with practice, and that (2) this learning is specific to backward masking. [Work supported by the James S. McDonnell Foundation and NIDCD.]

10:00–10:15 Break

10:15

5aPP9. Categorical perception of emotional speech. Bea de Gelder and Jean Vroomen (Dept. of Psych., Tilburg Univ., Tilburg, The Netherlands)

Emotions are communicated through the prosody of speech. Are such expressions perceived categorically or rather in a multidimensional space without clear foci? The former alternative is in line with longstanding views about the existence of some basic emotions. Recent reports present evidence that facial expressions are perceived categorically [de Gelder *et al.*, Cognit. Emotion (in press)]. To explore this issue for auditory speech, a continuum between happiness and fear was created by varying pitch excursion, pitch height, and duration of an utterance via PSOLA. Categoricity of perception for auditory expressions was examined in (a) an auditory-only presentation and (b) an auditory–visual condition of simultaneous presentation of the auditory stimuli and facial expressions corresponding to the either of the two expressions. In the auditory-only conditions a clear CP phenomenon is observed. In the bimodal conditions the visual information modulates the perception of the speech emotion. The latter phenomenon is not observed in a patient suffering from prosopagnosia. The findings raise the issue of a common biological basis for facial and vocal expression perception.

10:30

5aPP10. Phoneme detection in resyllabified words. Jean Vroomen and Bea de Gelder (Dept. of Psych., Tilburg Univ., Tilburg, The Netherlands)

There is evidence that the human listener takes, in continuous speech, the onset of a syllable as the onset of a spoken word [J. Vroomen and B. de Gelder, J. Exp. Psychol.: Human Percept. Performance (in press)]. However, this syllabic strategy would fail in resyllabified words like "my bike is," pronounced as /mai bai kis/. In the present study, a phoneme monitoring task is used to investigate whether resyllabified phonemes are more difficult to detect than their nonresyllabified counterpart. The results show that (i) nonresyllabified phonemes can be detected faster than resyllabified phonemes, and (ii) that this effect is bigger for words that are not unique at their offset. This suggests that, in particular, late-unique resyllabified words should be more difficult to perceive, and it lends further credence to the idea that a syllable onset is taken as the onset of a word.

10:45

5aPP11. Estimation of nonsense syllable intelligibility. Christine M. Rankovic (Dept. of Speech Lang. Pathol./Audiol., Northeastern Univ., Boston, MA 02115)

An intelligibility estimation test comprising nonsense syllables that were recorded separately, and then concatenated, is under development for hearing aid assessment. Consonant-vowel-consonant syllables (CVCs) are presented along with a visual display on a computer monitor that identifies each syllable with an animated pointer as it is presented to the listener. After presentation of a list of 50–60 CVCs, the listener is asked to estimate the consonant percent-correct score. A major advantage of the test is rapid administration. Percent-correct estimates were collected from 16 normal-hearing subjects presented with CVCs masked by a speech-shaped noise and with high- and low-pass filtered CVCs. In all test conditions, actual percent-correct scores were obtained for comparison. In noise and under filtering, estimates are slightly higher than actual scores. However, estimates increase with speech-to-noise ratio (dB) at a rate of about 4% per dB, similar to the actual percent-correct scores. The crossover frequency derived from the filtering experiment was approximately 1700 Hz, nearly identical to that reported by French and Steinberg [J. Acoust. Soc. Am. **19**, 90–119 (1947)]. Hence, articulation theory principles may be applicable. This test may serve as a speedy assessment tool for evaluating and selecting hearing aids. [Work supported by NIH.]

11:00

5aPP12. Calculation of speech intelligibility using an autocorrelation function. T. Shoda (Graduate School of Sci. and Technol., Kobe Univ., Rokkodai, Nada, Kobe, 657 Japan) and Y. Ando (Kobe Univ., Kobe, 657 Japan)

In this paper, a new method to calculate Japanese CV monosyllable speech intelligibility (confusion rate) by use of an autocorrelation function (ACF) is proposed. This method is based on the principle of speech information processing in the auditory pathway and the brain. In order to find out the essential part to identify syllables, the head and tail parts of syllables were cut and speech intelligibility tests were conducted. The normalized ACF of the essential part was analyzed. Results show that all syllables may be characterized by the time (τ) and the amplitude (ϕ) of only the first dip of the normalized ACF. Using this characterization, together with the distance between syllables and the manner of articulation, the monosyllable speech confusion rate may be calculated.

11:15

5aPP13. Feature extraction of phonemes from autocorrelation functions of auditory filterbank outputs. Koichi Sato, Jun Toyama, and Masaru Shimbo (Dept. of Information Eng., Faculty of Eng., Hokkaido Univ., Kita-ku, Sapporo, 060 Japan)

Features extracted from auditory filterbank outputs are useful as a front end for speech recognition systems. If an auditory filter has a center frequency (CF)[Hz], the extracted feature has been, in most cases, used at the CF[Hz] itself. In the case of inputting an f [Hz] pure tone into the auditory filterbank, the feature spreads across a wide frequency, though the output frequencies of any filters are f [Hz]. Therefore, the feature must be used not at CF[Hz], but at a special frequency which depends on the input wave. An autocorrelation function is useful for extracting such a special frequency. If the function has peaks at every T [s], the output of the filter mostly includes the component of $1/T$ [Hz]. Thus the special frequency and the feature can be extracted from the peaks of the autocorrelation function. As a result, the features extracted from all filters are not spread across a wide range of frequency. At low frequencies (less than about 1 kHz), the features are distributed among the pitch of the voice and its harmonies. At high frequencies, where pitch harmonies are not separated, the features are distributed among voice formants.

11:30

5aPP14. Application of frequency modulation technology to children. Barbara Franklin (Dept. of Special Education, San Francisco State Univ., 1600 Holloway Ave., San Francisco, CA 94132)

Classrooms provide a poor acoustical environment for children who have a hearing loss. Frequency modulation (FM) systems can reduce the signal-to-noise problems of hearing aids by providing a constant sound pressure of the speaker's voice. A new type of FM receiver combines the hearing aid and the FM system in a single behind-the-ear unit (BTE/FM). This new BTE/FM system eliminates the body-worn case as well as all loops and cords. A new product, which will shortly be introduced into the market, is an FM system, where all the components are contained in the "boot" which is then attached to the bottom of the hearing case. In this way, the hearing aid itself converts into an FM system. This paper will discuss this new technology and its contribution to children with a combined hearing and vision loss. [Work supported by OSERS.]

11:45

5aPP15. Study on detection of the electronic siren from an ambulance. H. Baba (Dept. of Transport Mech. Eng., Kurume Inst. of Technol., Kamitsu Kurume, 830 Japan) and M. Ebata (Kumamoto Univ., Kurokami Kumamoto, 2-39-1 Japan)

The electronic siren used on an ambulance is very loud in the street, but it is hard for the driver in an automobile to recognize it in many cases. It is conceivable that the reason is due to masking of the sounds inside and outside the automobile. In this paper, the detectable level of the electronic siren used on an ambulance was described and the reason why the thershold level of the siren is actually increased is discussed. As a result, conclusions were extracted as follows. (1) The threshold of the siren increases at the rate of 1.0–1.7 dB, when the level of music listened to in an automobile is increased by 2 dB in L_{Aeq}. This is considered to be caused by masking. (2) In an automobile running at medium speed, the detectable level of the siren increases by about 0.5 dB compared to the case of standstill. This is concerned with attention to driving.